TECHNICAL REPORT

APPROVED: 01 February 2018

doi:10.2903/sp.efsa.2018.EN-1378

Evaluation of the data on clothianidin, imidacloprid and thiamethoxam for the updated risk assessment to bees for seed treatments and granules in the EU

European Food Safety Authority

Abstract

The European Commission has requested the European Food Safety Authority (EFSA) to perform an updated risk assessment as regards the risk to bees from the uses of the three neonicotinoid pesticides active substances clothianidin, imidacloprid and thiamethoxam applied as seed treatments and granules. In performing this evaluation, in accordance with Article 21 of Regulation (EC) No 1107/2009 and considering recital 16 of Regulation (EU) No 485/2013; EFSA has been asked to undertake a review of the new data relevant to the uses under consideration taking into account in particular, the new relevant data collected in the framework of the open call for data organized by EFSA in 2015 and any other new data from studies, research and monitoring activities. EFSA has established a specific methodology for the evaluation of the available data. A full description of such methodology and the results of its application are reported.

Key words: clothianidin, thiamethoxam, imidacloprid, bee, review, assessment methodology

Question number: EFSA-Q-2017-00677

Correspondence: pesticides.peerreview@efsa.europa.eu

Suggested citation: EFSA (European Food Safety Authority), 2018. Evaluation of the data on clothianidin, imidacloprid and thiamethoxam for the updated risk assessment to bees for seed treatments and granules in the EU. EFSA supporting publication 2018:EN-1378. 31 pp. doi:10.2903/sp.efsa.2018.EN-1378

ISSN: 2397-8325

Summary

The European Commission has requested the European Food Safety Authority (EFSA) to perform an updated risk assessment as regards the risk to bees from the uses of the three neonicotinoid pesticides active substances clothianidin, imidacloprid and thiamethoxam applied as seed treatments and granules. In performing this evaluation, in accordance with Article 21 of Regulation (EC) No 1107/2009 and considering recital 16 of Regulation (EU) No 485/2013; EFSA has been asked to undertake a review of the new data relevant to the uses as under consideration taking into account, among others, all the new data collected in the framework of the open call for data organised by EFSA in 2015 and any other new data from studies, research and monitoring activities.

Commission Implementing Regulation (EU) No 485/2013 amended the conditions for approval of the three active substances, all belonging to the group of neonicotinoids, for use in plant protection products. In particular the uses as seed treatment and soil treatment of plant protection products containing clothianidin, thiamethoxam or imidacloprid should be prohibited for crops attractive to bees and for cereals except for uses in greenhouses and for winter cereals. The measures were taken following the previous EFSA assessments of the risk to bees from these active substances finalised in July 2013. In accordance with recital 16 of Regulation (EU) No 485/2013, within two years from the date of entry into force of that Regulation, the European Commission foresees to initiate without undue delay a review of the new scientific information available. For this purpose, following the European Commission request, on 25 May 2015 EFSA launched an open call for data. The outcome of the call, alongside with the complete list of contributions submitted, has been published by EFSA in the form of a Technical Report in November 2015.

On 16 November 2015, EFSA received a second follow-up mandate by the European Commission requesting to perform an updated risk assessment as regards the risk to bees from the uses of the three neonicotinoid pesticides active substances clothianidin, imidacloprid and thiamethoxam applied as seed treatments and granules, taking into account in particular the new data collected from the open call and any other new data from studies, research and monitoring activities that are relevant for the uses under consideration.

To address the mandate, EFSA also considered the data available from a previous systematic literature review, outsourced in 2013. Furthermore, an update of this systematic review was performed in June 2016, in order to collect all published scientific literature relevant for the current evaluation.

This technical report illustrates the specific methodology that has been designed by EFSA in order to obtain a structured and documented approach relating to the evaluation of the available data for what concern their relevance for the current risk assessment and their scientific reliability. Furthermore, the results of all the steps of such evaluation are also reported in this technical report.

Table of contents

1. Introduction

1.1. Background and Terms of Reference

Commission Implementing Regulation (EU) No 485/2013[1] amended the conditions for approval of the active substances clothianidin, imidacloprid and thiamethoxam for use in plant protection products, all belonging to the group of neonicotinoids. The specific provisions of the approval were amended to restrict the uses of clothianidin, thiamethoxam and imidacloprid, in order to provide for specific risk mitigation measures for the protection of bees and to limit the use of the plant protection products containing these active substances to professional users. In particular the uses as seed treatment and soil treatment of plant protection products containing clothianidin, thiamethoxam or imidacloprid should be prohibited for crops attractive to bees and for cereals, except for uses in greenhouses and for winter cereals.

The measures were taken following the previous EFSA assessments of the risk to bees from these active substances (EFSA, 2013a, 2013b, 2013c). In accordance with recital 16 of Regulation (EU) No 485/2013, within two years from the date of entry into force of that Regulation, the European Commission foresees to initiate without undue delay a review of the new scientific information received.

For this purpose, EFSA has been requested by a first mandate from the European Commission received in February 2015 to organise an open call for data for new scientific information as regards the risk to bees from the neonicotinoids clothianidin, thiamethoxam and imidacloprid in accordance with Article 21 of Regulation (EC) No 1107/2009[2] and in the context of Article 31 of Regulation (EC) No 178/2002[3]. A second mandate has subsequently been received by EFSA in November 2015, requesting to perform in accordance with Article 21 of Regulation (EC) No 1107/2009 and considering recital 16 of Regulation (EU) No 485/2013, an updated risk assessment as regards the risk to bees from the uses of the three neonicotinoids applied as seed treatments and granules. In performing this task EFSA should take into account in particular the new relevant data collected in the framework of the open call for data and any other new data from studies, research and monitoring activities that are relevant to the uses under considerations.

1.2. Interpretation of the Terms of Reference

The open call for data was launched on 25 May 2015 and closed on 30 September 2015. A technical report containing all data submitted has been prepared by EFSA and published on 13 November 2015 (EFSA, 2015a).

In order to take into account any new relevant data, EFSA also considered the data available in a previous systematic literature review on the three neonicotinoids and the risks to bees, outsourced in 2013 (Fryday et al., 2015) and performed an update of this systematic review in June 2016, in order to collect all published scientific literature relevant for the current evaluation.

Furthermore, Member State were requested by EFSA to provide additional data concerning the monitoring activities available at national level in relation to the specific uses under consideration and two Member States (Italy and Sweden) provided their feedbacks; which were also considered during the evaluation.

Finally, EFSA liaised with Member States and producers in order to investigate the uses they would like to support for the EU market (GAP tables).

[1] Commission Implementing Regulation (EU) No 485/2013 of 24 May 2013 amending Implementing Regulation (EU) No 540/2011, as regards the conditions of approval of the active substances clothianidin, thiamethoxam and imidacloprid, and prohibiting the use and sale of seeds treated with plant protection products containing those active substances, OJ L 139, 25.5.2013, p. 12–26

[2] Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC, OJ L 309, 24.11.2009, p. 1–50

[3] Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety, OJ L 31, 1.2.2002, p. 1–24

As a second step, EFSA has developed a specific methodology that was used for the evaluation of all the available scientific information. The methodology delineated in this technical report describes a structured, documented approach spanning over the whole process, from the data collection to the endpoints evaluation, in order to accomplish an impartial assessment of the available studies.

The methodology described in this technical report was employed by EFSA for evaluating all the data collected for the review of the risk assessment to bees from the three neonicotinoids used as seed treatment and granules. The results of the full evaluation of the data are also reported.

2. Description of the methodology

2.1. Source of information

Three main sources were used to gather all documents used in the review process.

The first source of data was **the open call for data** for new scientific information as regards the risk to bees from the use of the three neonicotinoid pesticide active substances clothianidin, imidacloprid and thiamethoxam applied as seed treatments and granules in the EU. The open call was launched in May 2015, upon request of the first mandate received by the EU Commission, and closed in September 2015. More details on the open call for data and the full list of contributions collected are available in a dedicated technical report of EFSA (EFSA, 2015a).

The second source of data was the **systematic literature search on the neonicotinoids and the risks to bees** that EFSA outsourced in 2013. This systematic search gathered publications on imidacloprid, thiamethoxam, and clothianidin from more than ten different databases, from the publication year of 1990 to October 2014. The aim was to collect all papers reporting any relevant information regarding potential exposure and effects to bees. The outcome of this literature search is reported in detail in a dedicated report (Fryday et al., 2015). In the aforementioned report, the authors also performed a screening based on a few simple criteria, in order to dismiss clearly irrelevant documents. Only the documents passing such screening were considered in the present review process.

Finally, In order to capture relevant information published after October 2014, an **update of the systematic literature search** was performed by EFSA. For the sake of consistency, the search strategy from the previous systematic search was maintained. However, due to access limitations, fewer databases were mined in the update. Nevertheless, it was considered that the searched archives (7 databases hosted under Web of Science) provided a sufficient coverage for quality published peer-review literature. The literature search was performed on 13 June 2016, therefore literature published after this date was not captured in the systematic literature search and not considered in the following steps of the evaluation.

2.2. Structure of the assessment workflow

The assessment procedure was implemented in a nested design, where the scale of the assessment changed at every step.

In the first step, which was the **title/abstract screening**, the focus of the assessment was on the whole document (e.g. a paper, a report, etc.), with the aim of assessing whether it contained any kind of information that may potentially inform the risk assessment.

Following the screening of titles and abstracts the second step was the **full text screening**. Often the same document could report information for several experiments (e.g. different experimental set-ups, effects investigated for different routes of exposure, different test locations, different time periods, etc.). When this was the case, each experiment was assessed individually to check once again whether each experiment contained any kind of information that may potentially inform the risk assessment. In parallel with the full text screening, each experiment was classified according to the appropriate experiment category; a more detailed description of the experiment categories is provided in Section 2.5.

Finally, within each experiment, there might be measurement of several response variables/outcomes, which in this report are generically identified with the term 'endpoint'. Each of these endpoints was assessed individually for its relevance and its reliability in the last step of the process, which combined the **data extraction** and the **critical appraisal**.

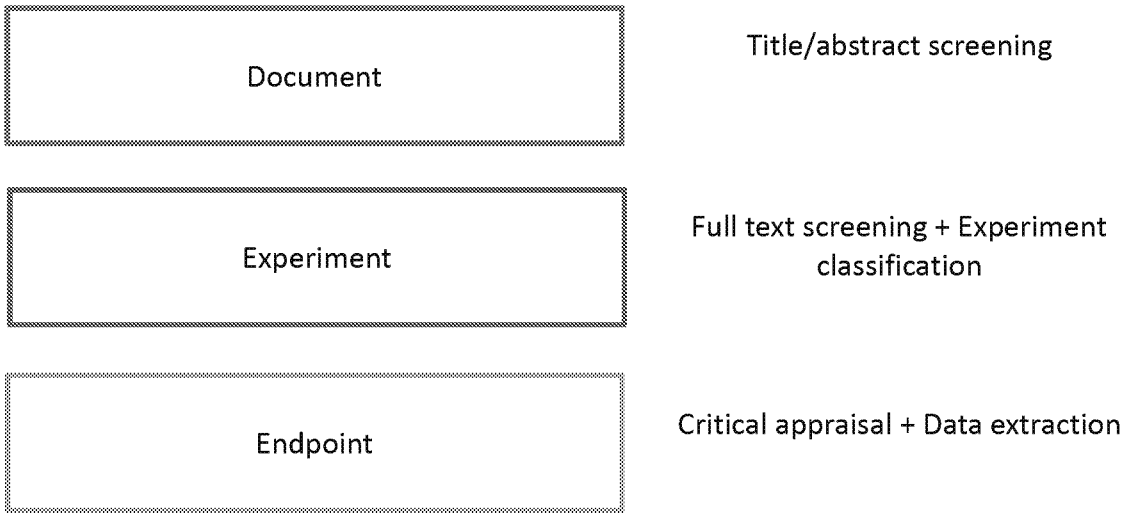

Figure 1: Schematic summary of the structure of the assessment workflow

2.3. Title/abstract screening

The screening was the very first step of the evaluation and was based on the principles of the systematic literature search methodologies as also reported in the respective phase described in the systematic literature review report on neonicotinoids and the risks to bees (Fryday et al., 2015).

The main scope of the title/abstract screening phase was to discriminate those studies that are likely to provide useful information from those that are obviously not useful for the current risk assessment. In practice, the purpose of this screening was to dismiss all those studies that are clearly not related to the topic of interest. Studies excluded during the title/abstract screening phase did not undergo a full assessment.

The title/abstract screening was based on only two alternative eligibility criteria, which were reported to the reviewer in terms of questions.

1. Does the document contain any information/data on the [potential] exposure of bees in accordance with EFSA (2013d)?

   AND/OR

2. Does the document contain any information/data on the [potential] effects that exposure to at least one of the three substances under assessment (including mixtures, formulations and metabolites) causes to any bee species?

Fulfilment of at least one of the aforementioned criteria was needed to pass the title/abstract screening. Due to extremely wide variety of documents, more specific conditions were adopted for particular topics. Documents focussing on analytical methods were only considered further when these methods were related to bee-relevant matrices (pollen, nectar, guttation fluid, dust, and dead bees) in accordance with the risk assessment scheme presented in EFSA (2013d). Documents focussing on the environmental fate of the chemicals under investigation were further considered only when they reported information on soil (e.g. degradation/dissipation, sorption, etc.) and on the bee-relevant matrices mentioned before. Documents reporting monitored environmental concentrations were considered only when reporting concentrations in pollen, nectar, and guttation fluids.

During this phase, in case of uncertain decision about one question, the criterion was considered fulfilled and the document selected for the full text assessment. The title-abstract screening was operatively performed by three Reviewers which used a tailored excel-based tool. The complete dataset was split in approximately 3 equal parts, each screened by one Reviewer.

Each of the study passing the title/abstract screening was allocated to a different substance-specific **assessment stream**. Streams for combination of substances (with or without interaction) were also

foreseen. A specific stream was also considered for those documents reporting non-substance specific information that may potentially inform the exposure assessment of bees (e.g. weeds coverage, etc.). Whenever there was a doubt regarding the appropriate allocation to a stream, a second Reviewer was consulted.

2.4. Full text screening

All documents retained after the title/abstract screening passed to the second step, which was another screening, based on the full text. However, the focus of the full text screening was on the single experiment rather than on the whole document. Every experiment reported in a document was evaluated individually.

The two alternative eligibility criteria used for the title/abstract screening were also considered for the full text screening. However, one additional criterion was employed.

Overall, the following scheme was adopted:

1. Does the experiment contain any information/data on the [potential] exposure of bees in accordance with EFSA (2013d)?

   AND/OR

2. Does the experiment contain any information/data on the [potential] effects that exposure to at least one of the three substances under assessment (including mixtures, formulations and metabolites) causes to any bee species?

   AND

3. Is the experiment primary research?

This additional criterion was meant to exclude from the full assessment documents or parts of documents that are not reporting primary research. These include description of experiments previously reported elsewhere (e.g. reviews) and also statements that are not reporting any experiment (e.g. position/discussion papers, generic statements, news, etc.). This criterion was implemented in order to avoid considering several times the same evidence (i.e. as primary research evidence per se and as evidence considered to inform review papers).

This phase of the methodology, together with those described under Sections 4.3 and 4.4, was practically implemented using the software tool DistillerSR. Separate (but structurally analogous) DistillerSR projects were created for each assessment stream, allowing full traceability of the assessment process at any time point. Three couples of Reviewers were established, each formed by one ecotoxicologist and one expert of environmental fate and behaviour of chemicals. Each couple of Reviewers focussed on one of the three substances under assessment, and therefore mainly acting on the respective stream/DistillerSR project.

2.5. Experiment type classification

Simultaneously with the full text screening, the experiment types were identified. The experiment classification was implemented for several reasons. First, different data extraction and appraisal schemes needed to be developed for each experiment type, in order to be fit-for-purpose. Secondly, this classification would ease future retrieval of information from experiments/documents excluded in the full text screening phase.

Originally, 21 different experiment types were foreseen. However, due to the wide spectrum of data retrieved from the different sources, it was necessary to implement further experiment types, reaching 24 types (Table 1). The methodology and its implementation in DistillerSR were designed to be flexible for this purpose, so that new experiment types could be added at any time.

Table 1: Full list and description of experiment types

Category	Experiment type	Description
1.	**Laboratory effects**	Any experiment carried out in the lab, which is focused on measuring effects on bees. Most of the experiments measuring endpoints at individual level (e.g. HPG development) belong to this category. Experiment at sub-individual level; TK/TD; internal exposure are also considered to be part of this category.
2.	**Semi-field (tunnel) [seed treatment or granules, dust deposition**	Tunnel test or similar (e.g. OECD75), which is focused on measuring effects on bees. The test substance is applied as seed treatment, granules, directly to soil, or for simulating dust drift.
3.	**Effects on colony – exposure via feeder**	Experiment where effects to bees are commonly measured at the colony level (e.g. brood development). Exposure is predominantly/exclusively via contaminated food provided to the bees by means of a feeder.
4.	**Field [seed treatment or granules]**	Experiments carried out in the open field where effects on bee colonies are measured following a certain exposure to the investigated chemical, when this is applied as seed treatment, granules, directly to soil, or for simulating dust drift. Very often the exposure is characterised by measurement of residues: this part of the document should be considered as a separate experiment belonging to Type 5 below.
5.	**Residue measurement [soil applications]**	Measurement of residues in bee-relevant matrices, following a known (rate, time, etc.) application of the investigated chemical as seed treatment, granules, or soil application. This category also includes studies on dust drift (residue quantification in dust, deposition fraction) and residues found in guttation (while Type 8 below is referring only to guttation occurrence).
6.	**Bee monitoring [seed treatment or granules]**	Experiments where the health status of colonies is checked. Colonies are placed within a landscape where a relevant area percentage is treated with the investigated substance applied as seed treatment, granules, or directly to soil.
7.	**Weeds coverage**	Experiment where the weed coverage of at least one field crop is characterised.
8.	**Guttation occurrence**	Experiment where the occurrence (e.g. frequency in space and time) of guttation is characterised in field conditions.
9.	**Dustiness**	Experiment where the production of dust from treated seeds of treated seeds is quantified (e.g. Heubach value). Normally carried out in the lab.
10.	**Environmental fate**	Experiments focusing on the e-fate of one or more of the substances under assessment. At the screening phase it was decided to include only those experiments reporting information for establishing the fate in pollen, nectar, and guttation fluids. Also information on the soil dissipation and sorption of the substances was retained.
11.	**Avoidance**	Experiments where the avoidance/preference of bees towards foodstuff contaminated with the three substances are reported.
12.	**Monitoring data (environmental concentrations)**	Studies where residues/environmental concentrations of a substance are measured and reported. In this kind of studies, differently to those belonging to Type.5, there is no possibility to link residues/environmental concentrations to a specific use of a substance. At the screening step it was decided to retain only studies reporting this kind of data for pollen, nectar, guttation fluids; therefore concentrations found in soil and water/sediment were not extracted nor assessed.
13.	**Efficacy experiments**	Effects on organisms other than bees are unlikely to be relevant for the present risk assessment.
14.	**Agronomic practices**	Experiments of any kind which provide information on agricultural practice that may have an impact on the exposure assessment.
15.	**Analytical methods**	At the screening level it was decided to consider only those

Category	Experiment type	Description
		studies presenting methodologies for extracting and quantifying at least one of the three substances in: pollen, nectar, guttation fluids, dust, and dead bees
16.	**Comprehensive risk assessment statement/position papers**	This kind of document generally aims at giving a general overview of the existing data within the scope of the risk assessment. Normally, no new data are reported in these documents (not primary research). Note that in these cases very often no (new) experiments are actually reported in the document. Therefore within the present methodology it was considered that document=experiment.
17.	**General review**	This kind of document by definition does not provide new data (not primary research). Also in these cases, within the present methodology it was considered that document=experiment.
18.	**Semi-field (tunnel) [other applications]**	Experiments as those in type 2, but with different application techniques
19.	**Field [other applications]**	Experiments as those in type 4, but with different application techniques.
20.	**Residues measuring [other applications]**	Experiments as those in type 5, but with different application techniques.
21.	**Bee monitoring [other applications]**	Experiments as those in type 6, but with different application techniques.
22.	**Epidemiological analysis**	Analysis of epidemiological data assessing bee health in relation to several risk factors (not foreseen in the initial type list).
23.	**Palynological analysis**	Analysis of pollen types/sources, generally performed on pollen collected by bees (not foreseen in the initial type list).
24.	**Metabolism study with bees**	Experiments investigating the metabolism of a substance in bees.
25.	**Others**	Any other experiment type, not classifiable in any of the previous types (not foreseen in the initial type list).

Experiments from different types were treated differently, in accordance with their relevance for the risk assessment. Three groups were identified.

- The **first group** included the experiment types that were considered to provide the most significant information to the risk assessment. Endpoints reported in these experiments were extracted and underwent a full assessment. Experiments types 1-6 were included in this first group, as according to previous experience in this kind of risk assessment, these experiment types are likely to provide the most pivotal data.
- The **second group** included experiment types that might provide useful information, but for which assessment guidance is lacking in the context of the current assessment. Endpoints were extracted, but their assessment was carried out in a more flexible manner compared to group 1. This group included experiment types 7-11. Data from group 23 were not extracted as such, but were considered for interpreting effect studies.
- The **third group** included experiment types that, for the time being, are considered only marginally relevant for the present assessment. Endpoints from these experiments were not extracted or assessed. This group included experiment types 12-25, for different reasons: lack of fit with the present risk assessment scheme (types 12-15, 22, 24, and 25); secondary results only (types 16-17); non-relevant application techniques (types 18-21).

2.6. Data extraction

Entirely quantitative endpoints (such as NOEC/LD50 from dose-response studies, residue concentrations, etc.) were extracted as such. For effect endpoints typical of higher tier studies (e.g. brood production, brood index, colony strength, forager mortality, etc.) a more narrative description of the results is extracted. In this latter case, consideration was particularly given to the deviance between treatment(s) and control(s), accounting for the magnitude of this deviance, its duration, and the time trend in general.

For all endpoints basic metadata (e.g. tested organisms, substance tested/analysed, number of replicates, doses/dressing/application rates, study location etc.) were extracted as well.

2.7. Critical appraisal

2.7.1. Endpoint evaluation criteria

A specific list of criteria was compiled for the six experiment types belonging to the first group as described in Section 2.5, which are the study types considered as most important for the risk assessment. These criteria are largely based on the recommendations included in the EFSA guidance document (EFSA, 2013d); however, the CRED method (Moermond et al., 2015), despite being developed for aquatic ecotoxicity testing, was also considered in this context.

The assessment criteria, written in form of questions, are meant to be a guide for the Reviewer during the relevance and the reliability evaluation of the endpoints, in accordance with the scheme reported in Figure 1: .

It was considered that a simple yes/no answer to each criterion was not appropriate to represent the complexity of some issues. A sort of 'fulfilment scale' was considered more appropriate in this context. Therefore, each of the possible responses reflects the degree of fulfilment for a specific criterion.

Table 2: Description of the fulfilment scale used for the critical appraisal

Answer	Fulfilment scale explanation
Yes	Criterion fully satisfied
Partially	Most of the aspects considered in the criterion were satisfied
Barely	Only a few aspects considered in the criterion were satisfied
No	Criterion not satisfied
Not reported	No information is available to assess the fulfilment of the criterion
Not relevant	Criterion not relevant for the specific study under assessment

A different set of questions/criteria was drafted for each experiment type. However, a common structure of 7 sections was maintained across types.

1) **Relevance:** This set of questions should help the Reviewer deciding the relevance of each endpoint for the present risk assessment. This includes whether the measured endpoint fits within the risk assessment scheme presented in EFSA (2013d) at lower tiers and/or whether it provides information that could be quantitatively related to the protection goals. However, relevance also includes consideration about:

- exposure route
- pesticide application (e.g. technique, season, timing in relation to the GAP under assessment)
- study location (climate, geopedology in relation to the intended application area)
- crop (i.e. is the crop considered in the GAP? If not, a consideration should be given to the relative attractiveness and the crop-specific residue levels).

2) **Reliability-test organism:** This includes the identification of the tested organisms, information on their origin, the reporting of an appropriate initial health status, absence of previous stress, and appropriate handling methodology (acclimatisation, test conditions).

3) **Reliability-test setup:** These set of questions includes the aspects of the test design related to the experimental setting, considering the adherence to existing guidelines, the environmental conditions, the presence of controls, the sampling and observations of the outcome variables. It also includes the reporting quality about any other detail of the test setup, including the identification of the study location (except for laboratory studies), the

application technique and its rate/dosing (and the form/unit in which this is expressed), and the frequency and the duration of the sampling/observation.

4) **Reliability-exposure characterisation:** These set of questions considers the necessary information for characterising the exposure in terms of chemical identification (e.g. active substance vs. formulation, form, purity, content of active substance in a formulation, presence of potentially influential co-formulants), measurement, influential landscape, and environmental conditions. It also includes a clear reporting on the exposure duration, the availability of analytical measurement (residues or confirmed doses), information on test item consumption (for oral exposure), and any other relevant information on the fate of the tested item.

5) **Reliability-test design and statistics:** These set of questions includes the aspects of the test design more related to statistical issues and the post-experimental treatment of the data. This includes a clear identification on the response variables (endpoints), an appropriate sample/observation unit size (e.g. colony/cage size in terms of individuals), an appropriate replication, randomisation, and an appropriate number of tested doses and spacing between them (for dose-response tests). This also includes presenting the data with appropriate elaborations and correct statistic techniques. Finally, availability of raw data is also checked.

6) **Reliability- analytics:** This includes criteria for the assessment of analytics, where chemical analysis is part of a certain experiment. This section makes large use of the EU Draft Guidance for residues methods of analysis (European Commission, 2000).

7) **Reliability-conclusions:** This includes consideration about whether the conclusion of the experiment is supported by the data/results.

A full overview of the criteria employed for each of the six experiment types belonging to the first group is presented in Appendix A.

2.7.2. Endpoint evaluation outcome

Within a single experiment, several endpoints might be presented. Pending on the study design and on the data reporting, the relevance and the reliability level can differ between endpoints within the same experiment.

Relevance was considered with respect to the protection goals. Some endpoints were considered ***directly relevant*** either because they serve as input for the lower tier of the risk assessment (e.g. toxicity data, residue data, etc.) or because they are directly linked to the protection goals (e.g. colony strength and forager mortality for the honey bees). Other endpoints provide useful information, but they are only ***indirectly relevant*** for the risk assessment, as they don't fit with the scheme at lower tiers and their relationship with the protection goals cannot be quantified at the time of the assessment. Other endpoints were considered ***not relevant***, particularly when it is impossible to establish even a qualitative relationship with the assessment at the colony level (for both exposure and effects).

The quality assessment of the endpoints, from the experiment design to the reporting, is expressed as a **reliability** judgment. A 4-level rating scale was used, similar to the one reported in Klimisch et al. (1997). Studies were classified as:

1. Fully reliable
2. Reliable with minor restrictions

3. Reliable with major restrictions
4. Not reliable/Not assessable

It should be noted that Klimisch and colleagues gave a great weight to good laboratory practice (GLP) compliance for the reliability assessment, while in the present methodology GLP compliance was not considered as a guarantee of reliability, as recommended in EFSA (2011).

Reliability and relevance of all available endpoints were further considered in the final risk assessment, when homogenous results from all the available experiments were integrated.

3. Testing phase (ring test)

A ring test was performed before drafting the final version of the evaluation methodology, which is reported in the present technical report. The aim of the ring test was to fine tune the assessment methodology, highlighting the most critical (subjective) phases of the process, and therefore to identify where to invest more energy in clarifying and harmonising. It has to be stressed that the results of this ring test were not obtained with the current methodology, which was amended precisely on the basis of the outcome of this ring test. However, for sake of transparency, the methodology and the results of the ring test are briefly reported here.

A group of five Reviewers was involved in the ring test. All of these Reviewers have extensive experience (> 5 years) in environmental risk assessment of pesticides. However, two Reviewers had more expertise in environmental fate and behaviour (exposure assessment), while the remaining three had more expertise in ecotoxicology (effect assessment). Each of the Reviewers was tasked with assessing the same 7 documents chosen by another experienced scientist. Each Reviewer acted independently from the others.

The consistency of the assessment was analysed at different steps, in accordance with the phases of the evaluation methodology. At all stages, the agreement among the Reviewers was represented by a score normalised between 0% (maximum possible disagreement, maximal deviation among the answers) and 100% (maximum possible agreement).

The score for the screening phase was given by the inclusion/exclusion of the documents on the basis of the eligibility criteria (Table 3).

The consistency of the identification of the experiment types was assessed by measuring the fraction of the Reviewers that identified a certain experiment type within a certain document (Table 4).

The following phases of the process are based on each single endpoint. However, as not all Reviewers always identified the same endpoints, here are presented only the results for the endpoints identified by at least 3 Reviewers. Consistency was assessed for the responses (level of fulfilment) to each of the criteria in the appraisal process (full assessment) and for the final judgment assigned for relevance and reliability.

The results for the screening phase (inclusion/exclusion of documents on the basis of the eligibility criteria) are presented in Table 3. In this section, results are presented for the aggregation of the title/abstract screening and the full text screening. A total agreement was achieved for 5 out of the 7 documents. However, some disagreement was present for the remaining 2 documents, particularly for document n.3, which was included by 3 reviewers and excluded by 2.

Table 3: Results of the ring test - screening phase. The columns "Included" and "Excluded" report the number of Reviewers that concluded accordingly for each document.

Document n.	Included	Excluded	Agreement
1	4	1	50.0%
2	0	5	100.0%
3	3	2	0.0%
4	5	0	100.0%
5	5	0	100.0%
6	5	0	100.0%
7	5	0	100.0%
	Average		78.6%

The identification of the experiment types caused some disagreement among the Reviewers. However, it has to be highlighted that at the time the ring test was carried out, clear definitions for each experiment type (as those presented in Section 2.5) were not yet available. Anyway, even before this fundamental amendment, the agreement level was already around 70%.

Table 4: Results of the ring test – identification of the experiment types. The numbers in the table represent the fraction of the Reviewers that identified a certain experiment type within a certain document. The numbers in brackets are the associated agreement score, which is averaged by document in the last column.

Document n.	Type 1	Type 3	Type 4	Type 5	Type 20	Type 6	Others	Average
Doc 1	0.6 (0.5)	1.0 (1.0)						0.75
Doc 3					0.2 (0.0)	0.2 (0.0)	0.4 (0.25)	0.08
Doc 4		1.0 (1.0)						1.0
Doc 5			1.0 (1.0)	0.8 (0.75)				0.87
Doc 6			0.4 (0.25)	0.6 (0.5)				0.37
Doc 7	1.0 (1.0)							1.0
							Total average	**0.68**

Overall, 11 endpoints were identified by at least 3 Reviewers. Agreement for the full assessment (level of fulfilment for each of the criteria in the appraisal exercise) was rather high. None of the appraised endpoint resulted in an agreement <50% (Figure 2:). Overall, the mean (and median) agreement level across endpoints was 70%.

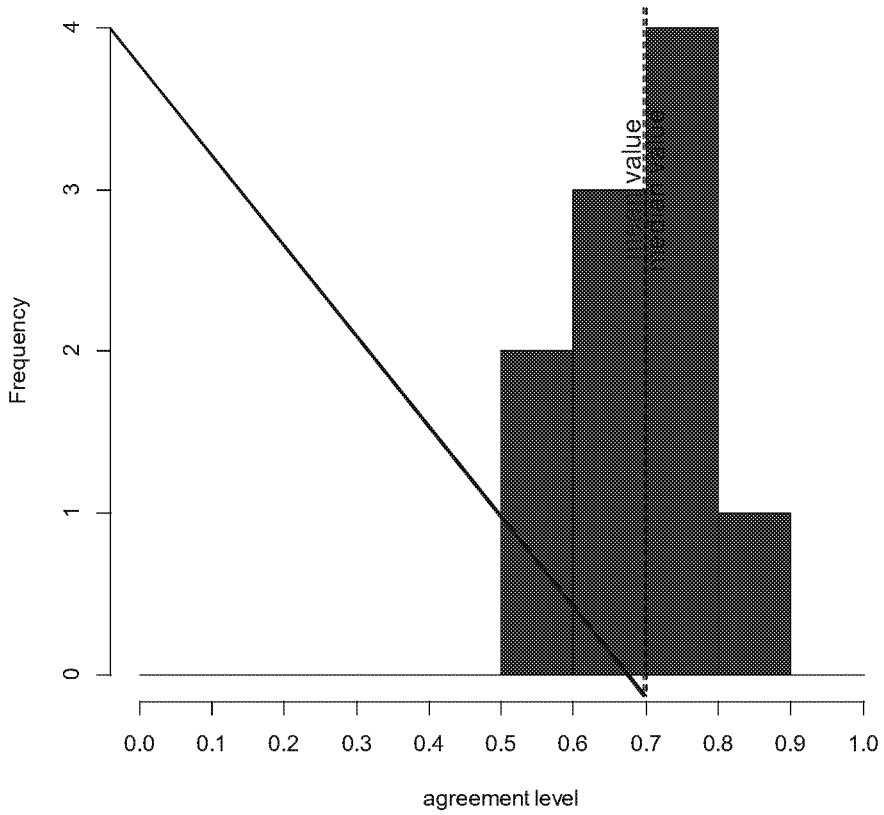

Figure 2: Frequency distribution of the agreement level among Reviewers for the full appraisal exercise. Only endpoints assessed by at least 3 Reviewers are considered.

The average agreement for the final relevance and reliability judgment were 63% and 60%, respectively. For sake of comparison, an endpoint considered 'Fully reliable' by two Reviewers and 'Reliable with minor restriction' by a third Reviewer, would result in an agreement level of 66.6%.

Overall, the ring test revealed that, thorough the whole process, the level of agreement among the Reviewers was between 60-80%. This was considered sufficient. However, some further measures were undertaken to improve the level of harmonisation and to reduce the risk for subjective interpretation. These included a clearer definition of the eligibility criteria in the screening, and a more formal definition of the experiment types. Finally, each of the Reviewer provided feedback on the aspects that were considered less clear. These were thoroughly considered and discussed in depth among the Reviewers, before drafting the final version of the current evaluation methodology. These further measures should have improved the quality of the assessment methodology, by reducing the subjectivity of the assessment and by harmonising the interpretation of the criteria.

4. Results

4.1. Retrieved literature

Overall, the first systematic literature search comprised 546 (already screened) documents, while the update of the literature search retrieved 874 documents. Finally, 376 contributions were received during the open call for data. After duplicate removal, the overall initial list included 1599 documents.

4.2. Title/abstract screening

The documents were randomly allocated to three Reviewers. Overall, they acted with a rather consistent exclusion rate (range 50-57%, average 54%, CV= 7%), proving that the influence of the Reviewer in this phase was limited.

During this phase, further duplicate documents were identified. In the updated literature search, some EFSA outputs and the report of the previous literature search (Fryday et al., 2015) were also retrieved. These documents were obviously not further considered.

After the abstract/title based screening, the dataset was reduced to 735 references. For all of these, an attempt to retrieve the full text was made.

During the full text retrieval, other 24 duplicates were identified. Other 12 documents were discarded because written in non-European languages (3 Japanese, 8 Chinese, 1 Russian). Finally, it was found that for 5 references only the abstract existed; these references were therefore excluded. Full text was not found for 14 references (2% of the amount potentially considerable for the review process). Each of the retrieved full text was allocated to a specific assessment stream (Table 5:).

Table 5: Number of retrieved full text allocated to each stream

Assessment Stream	Number of documents
All* (with interaction)	3
All+ (no interaction)	108
Clothianidin	111
Imidacloprid	259
Thiamethoxam	69
Clothianidin * Imidacloprid (with interaction)	8
Clothianidin + Imidacloprid (no interaction)	32
Clothianidin * Thiamethoxam (with interaction)	8
Clothianidin + Thiamethoxam (no interaction)	16
Imidacloprid * Thiamethoxam (with interaction)	2
Imidacloprid + Thiamethoxam (no interaction)	24
Not substance-specific	40

A full report of the title/abstract screening phase is reported in the Appendix B, together with the outcome of the full text retrieval process.

4.3. Full text screening / Experiment type classification

The number of included and excluded experiments per type at the level of full text screening is reported below (Table 6). The full report of this phase is reported in Appendix C to this Technical Report.

Table 6: Resume of the Full text screening / Experiment type classification phase.

Study type	Excluded	Included	Sum
Group 1 (Full appraisal and extraction performed, if included)			
1 - Laboratory effect	37	277	314
2 - Semi-field (tunnel) [seed treatment or granules, dust deposition]	2	25	27

Study type	Excluded	Included	Sum
3 - Effect on colony – exposure via feeder	9	65	74
4 - Field [application as seed treatment or granules]	26	75	101
5 - Residue measurement [soil applications]	17	99	116
6 - Bee monitoring [application as seed treatment or granules]	2	3	5
Total for group 1	***93***	***544***	***637***
Group 2 (Flexible appraisal and extraction performed, if included)			
7 - Weeds coverage	0	2	2
8 - Guttation occurrence	2	13	15
9 - Dustiness	4	7	11
10 - Environmental fate	26	12	38
11 - Avoidance	1	3	4
23 - Palynology analysis (not properly extracted)	0	7	7
Total for group 2	***33***	***44***	***77***
Group 3 (No appraisal or extraction performed)			
12 - Monitoring data (environmental concentrations)	27	24	51
13 - Efficacy experiments	1	0	1
14 - Agronomic practices	3	4	7
15 - Analytical methods	22	2	24
16 - Comprehensive risk assessment statement/position paper	22	0	22
17 - General review	45	0	45
18 - Semi-field (tunnel) [other applications]	3	13	16
19 - Field [other applications]	4	9	13
20 - Residue measurement [other applications]	6	6	12
21 - Bee monitoring [other applications]	11	3	14
22 - Epidemiological analysis	1	2	3
23 – Metabolism studies with bees	0	2	2
24 - Others	27	17	44
Total for group 3	***172***	***82***	***254***
Overall Total	**298**	**670**	**968**

4.4. Critical appraisal

A full appraisal was carried out for 588 experiments, belonging to the types of group 1 and 2 (see section 2.5). The results of the critical appraisal exercise are reported in the study evaluation notes reported in Appendices from D to O of this Technical Report.

5. Conclusions

The European Commission has requested the European Food Safety Authority (EFSA) to perform an updated risk assessment as regards the risk to bees from the uses of the three neonicotinoid pesticides active substances clothianidin, imidacloprid and thiamethoxam applied as seed treatments and granules. In performing this evaluation, in accordance with Article 21 of Regulation (EC) No 1107/2009 and considering recital 16 of Regulation (EU) No 485/2013, EFSA has been asked to undertake a review of the new data relevant to the uses as seed treatments and granules.

All the new data collected in the framework of the open call for data organised by EFSA in 2015 (EFSA, 2015a) are part of the information collected for this purpose.

In order to take into account any new relevant data, EFSA also considered the data available in a previous systematic literature review outsourced in 2013 (Fryday et al., 2015) and performed an update in June 2016, in order to collect all published scientific literature relevant for the current evaluation. EFSA has developed a specific methodology covering the whole process from the data collection to the endpoints assessment. Generic eligibility criteria were defined in the two screening phases, while other experiment type-specific criteria were drafted for the assessment of the endpoints.

The methodology was tested within EFSA in a preliminary version. The results of the testing phase showed that the designed methodology was sufficiently unbiased. Nevertheless, some amendments were carried out, in order to further strengthen the robustness and the clarity of the assessment methodology.

The methodology is published in the present technical report together with the results of its application in order to transparently show how the assessment of the available data was performed.

References

EFSA (European Food Safety Authority), 2011. Submission of scientific peer-reviewed open literature for the approval of pesticide active substances under Regulation (EC) No 1107/2009. The EFSA Journal 2011;9(2):2092 49 pp. doi: 10.2903/j.efsa.2011.2092. Available online: www.efsa.europa.eu

EFSA (European Food Safety Authority), 2013a. Conclusion on the peer review of the pesticide risk assessment for bees for the active substance clothianidin. The EFSA Journal 2013;11(1):3066 58 pp. doi: 10.2903/j.efsa.2013.3066. Available online: www.efsa.europa.eu

EFSA (European Food Safety Authority), 2013b. Conclusion on the peer review of the pesticide risk assessment for bees for the active substance imidacloprid. The EFSA Journal 2013;11(1):3068 55 pp. doi: 10.2903/j.efsa.2013.3068. Available online: www.efsa.europa.eu

EFSA (European Food Safety Authority), 2013c. Conclusion on the peer review of the pesticide risk assessment for bees for the active substance thiamethoxam. The EFSA Journal 2013;11(1):3067 68 pp. doi: 10.2903/j.efsa.2013.3067. Available online: www.efsa.europa.eu

EFSA (European Food Safety Authority), 2013d. Guidance on the risk assessment of plant protection products on bees (Apis mellifera, Bombus spp. and solitary bees). The EFSA Journal 2013;11(7):3295 266 pp. doi: 10.2903/j.efsa.2013.3295.

EFSA (European Food Safety Authority), 2015a. Technical report on the open call for new scientific information as regards the risk to bees from the use of the three neonicotinoid pesticide active substances clothianidin, imidacloprid and thiamethoxam applied as seed treatments and granules in the EU. EFSA supporting publication 2015:EN-903. 8 pp.

European Commission, 2000. Residues: guidance for generating and reporting methods of analysis in support of pre-registration data requirements for Annex II (Part A, Section 4) and Annex III (Part A, Section 5) of Directive 91/414. SANCO/3029/99-rev. 4, 11 July 2000

Fryday S, Tiede K and Stein J, 2015. Scientific services to support EFSA systematic reviews: Lot 5 Systematic literature review on the neonicotinoids (namely active substances clothianidin, thiamethoxam and imidacloprid) and the risks to bees. EFSA supporting publication 2015:EN-756, 656 pp.

Klimisch HJ, Andreae M, Tillmann U, 1997. A Systematic Approach for Evaluating the Quality of Experimental Toxicological and Ecotoxicological Data. Regulatory toxicology and pharmacology 25:1-5.

Moermond CTA, Kase R, Korkaric M, Ågerstrand M, 2016. CRED: Criteria for reporting and evaluating ecotoxicity data. Environmental Toxicology and Chemistry, 35: 1297–1309. doi: 10.1002/etc.3259

OECD, 2007. Series on testing and assessment number 75. Guidance Document on the Honey Bee (*Apis Mellifera* L.) Brood Test Under Semi-field Conditions

Abbreviations

CRED	Criteria for Reporting and Evaluating Ecotoxicity Data
CV	Coefficient of Variation
EU	European Union
GAP	Good Agricultural Practices
HPG	Hypopharyngeal gland
LD50	Lethal Dose for 50% of individuals
MS	Member State
NOEC	No Observed Effect Concentration
OECD	Organisation for Economic Co-operation and Development
TK/TD	Toxicokinetic/Toxicodynamic
GLP	Good Laboratory Practices

Appendix A – Overview of the relevance and reliability criteria employed for the six experiment types belonging to the first group

a. Experiment type 1 – Laboratory effects

Relevance
Could the endpoint be relevant for the risk assessment according to EFSA (2013)?
In case of mixture/formulation: is the test item representative and relevant for the a.s.?
Is the route of exposure relevant for the risk assessment?
Reliability-test organism
Is/are the tested organism(s) clearly identified (species level or lower)?
Are the life stages of the tested organism clearly identified?
Is the origin of the tested organism specified and trustable?
Were the test organisms properly acclimatised to the study setup before the exposure started?
Were the test organisms exempt from previous exposure or any kind of stressor?
Is the health status of the organisms checked and appropriate (are they disease-free)?
Is the handling of the organism appropriate? (to minimise any stress)
Reliability-test setup
Is the experimental part sufficiently detailed?
Is the study carried out according to a standard protocol/guidance?
If guideline/guidance is available, are validity criteria respected?
Is a negative control present?
Is a positive control present?
If no guideline/guidance is available for the study, is the negative control performing adequately?
If no guideline/guidance is available for the study, is the system sensitive?
If a solvent was used, is a solvent control present?
Is the application rate/dosing clearly reported?
Are the frequency and the duration of sampling/observation clearly identified and appropriate?
Is the colony (colonies) size appropriate for the study?
Is the study methodology consistent with the objective of the study?
Reliability- Exposure characterisation
Is the test item clearly identified and characterised?
For studies with technical - Is the purity of the a.s. reported?
For studies with formulations - Is the a.s. content specified?
For studies with formulations - Is the effect of coformulant investigated?
Is the exposure duration clearly indicated and appropriate?
Is exposure characterised by analytical measurements (residues or confirmed dose)?
Are the test conditions clearly reported (T, humidity, cages size and materials) and appropriate?
Is test item consumption measured/estimated and is it clearly reported?
Is the evaporation of the sucrose solution checked and accounted for?
Reliability- Test design and statistics
Are all measured endpoints/response variables clearly identified in the methodology description?

Is the test sample size/observation unit appropriate?
Is replication appropriate?
Is the number of tested concentration/doses appropriate?
Is a correct spacing adopted between tested doses/concentrations?
Are appropriate statistics available, described and consistent with the initial proposal of the study?
Are appropriate data elaboration presented?
Are raw data available?
Are units correct and consistent within the study?
Is the test design/sampling randomised?
Is an appropriate preliminary range-finding test performed?
Is a clear dose-response observed in the study?
Is the confidence interval of the endpoint reported and does it give an appropriate level of confidence on the derived endpoint?
Analytics
Is the analysed matrix clearly identified?
Is the sample storage and stability reported and appropriate?
Is equipment, materials and conditions used clearly described?
Is sample preparation techniques reported?
Is the method for the extraction and purification of the samples clearly described?
Is the analytical procedure clearly reported?
Is the described method appropriate for the intended analysis?
Linearity: is the analytical calibration extended over an appropriate range?
Linearity: is the number of determinations suitable for calibration?
Linearity: are details equation of the calibration line reported?
Linearity: is the overall calibration (linearity) satisfactory?
Accuracy: is mean recovery (+ RSD) reported for each investigated compound?
Accuracy: is mean recovery satisfactory?
Precision: is the RSD reported for each fortification level and for the overall recovery?
Precision: is the RSD < 20% per each fortification level?
LOQ: Is the LOQ clearly reported?
LOQ: are blank values (procedural blanks and untreated samples) < 30% of the LOQ?
Are the representative chromatograms reported?
Were the control samples analysed concurrently to determine any contamination by the analyte of interest or interferences?
Are all performed calculations clear and correct?
Is the analytical sample size appropriate?
Study conclusions
Is the conclusion of the experiment supported by the data/results?

b. Experiment type 2 – Semi-field (tunnel) [soil applications, dust deposition]

Relevance
Could the endpoint be relevant for the risk assessment according to EFSA (2013)?
In case of mixture/formulation: is the test item representative and relevant for the a.s.?
Is the application method relevant for the uses under evaluation?
Is the treated crop relevant for the investigated GAPs?
Is the study location relevant for the investigated GAPs?
Is the application period (season, growth stage) relevant for the investigated GAPs?
Reliability-test organisms
Is/are the tested organism(s) clearly identified (species level or lower)?
Is the origin of the tested organism specified and trustable?
Were the test organisms properly acclimatised to the study setup before the exposure started?
Were the test organisms exempt from previous exposure or any kind of stressor?
Is the health status of the organisms checked and appropriate (are they disease-free)?
Is the handling of the organism appropriate? (to minimise any stress)
Reliabiliy-test setup
Is the experimental part sufficiently detailed?
Is the study carried out according to a standard protocol/guidance?
If guideline/guidance is available, are validity criteria respected?
Is a negative control present?
Is a positive control present?
If no guideline/guidance is available for the study, is the negative control performing adequately?
If no guideline/guidance is available for the study, is the system sensitive?
Is the application rate/dosing clearly reported?
Is the application type clearly reported?
Is the application timing (date, season, crop growth stage) clearly reported?
Is the treated crop clearly identified?
Is the study location sufficiently identified in geographical terms?
Are the frequency and the duration of sampling/observation clearly identified and appropriate?
Is the sampling/observation technique reported and is it appropriate?
Is the colony (colonies) size appropriate for the study?
Are other crop management procedures (likely to have an impact) sufficiently detailed?
Are weather conditions sufficiently reported?
Is the plot unity of appropriate size?
Are conditions in treatments and control comparable at the beginnig of the experiment?
Is the soil typology identified (taxonomy, pH, OC%)?
Is the study methodology consistent with the objective of the study?
Reliability- Exposure characterisation
Is the test item clearly identified and characterised?
For studies with technical - Is the purity of the a.s. reported?
For studies with formulations - Is the a.s. content specified?

For studies with formulations - Is the effect of coformulant investigated?
Is the exposure duration clearly indicated and appropriate?
Is exposure characterised by analytical measurements (residues or confirmed dose)?
Is "dustiness" of treated seed reported (i.e. Heubach test)?
Reliability- Test design and statistics
Are all measured endpoints/response variables clearly identified in the methodology description?
Is the test sample size/observation unit appropriate?
Is replication appropriate?
Are appropriate statistics available, described and consistent with the initial proposal of the study?
Are appropriate data elaboration presented?
Are raw data available?
Are units correct and consistent within the study?
Is the test design/sampling randomised?
Is a clear dose-response observed in the study?
Study conclusions
Is the conclusion of the study supported by the data/results?

c. Experiment type 3 – Effects on colony – exposure via feeder

Relevance
Could the endpoint be relevant for the risk assessment according to EFSA (2013)?
In case of mixture/formulation: is the test item representative and relevant for the a.s.?
Reliability-test organisms
Is/are the tested organism(s) clearly identified (species level or lower)?
Are the life stages of the tested organism clearly identified?
Is the origin of the tested organism specified and trustable?
Were the test organisms properly acclimatised to the study setup before the exposure started?
Were the test organisms exempt from previous exposure or any kind of stressor?
Is the health status of the organisms checked and appropriate (are they disease-free)?
Is the handling of the organism appropriate? (to minimise any stress)
Reliabiliy-test setup
Is the experimental part sufficiently detailed?
Is the study carried out according to a standard protocol/guidance?
If guideline/guidance is available, are validity criteria respected?
Is a negative control present?
Is a positive control present?
If no guideline/guidance is available for the study, is the negative control performing adequately?
If no guideline/guidance is available for the study, is the system sensitive?
Is the dosing clearly reported?
Is the dosing timing (date, season, crop growth stage) clearly reported?
Is the study location sufficiently identified in geographical terms?
Are the frequency and the duration of sampling/observation clearly identified and appropriate?
Is the sampling/observation technique reported and is it appropriate?
Are landscape characteristics sufficiently described and appropriate (e.g. land use, potential dilution)?
Are other crop management procedures (likely to have an impact) sufficiently detailed?

Are weather conditions sufficiently reported?
Are conditions in treatments and control comparable at the beginnig of the experiment?
Is the study methodology consistent with the objective of the study?
Reliability- Exposure characterisation
Is the test item clearly identified and characterised?
For studies with technical - Is the purity of the a.s. reported?
For studies with formulations - Is the a.s. content specified?
Is the exposure duration clearly indicated and appropriate?
Is exposure characterised by analytical measurements (residues or confirmed dose)?
Is test item consumption measured/estimated and is it clearly reported??
Reliability- Test design and statistics
Are all measured endpoints/response variables clearly identified in the methodology description?
Is the test sample size/observation unit appropriate?
Is replication appropriate?
Is the number of tested concentration/doses appropriate?
Are appropriate statistics available, described and consistent with the initial proposal of the study?
Are appropriate data elaboration presented?
Are raw data available?
Are units correct and consistent within the study?
Is the test design/sampling randomised?
Is a clear dose-response observed in the study?
Study conclusions
Is the conclusion of the study supported by the data/results?

d. Experiment type 4 – Field [seed treatment or granules]

Relevance
Could the endpoint be relevant for the risk assessment according to EFSA (2013)?
In case of mixture/formulation: is the test item representative and relevant for the a.s.?
Is the application method relevant for the uses under evaluation?
Is the treated crop relevant for the investigated GAPs?
Is the study location relevant for the investigated GAPs?
Is the application period (season, growth stage) relevant for the investigated GAPs?
Reliability-test organisms
Is/are the tested organism(s) clearly identified (species level or lower)?
Is the origin of the tested organism specified and trustable?
Were the test organisms properly acclimatised to the study setup before the exposure started?
Were the test organisms exempt from previous exposure or any kind of stressor?
Is the health status of the organisms checked and appropriate (are they disease-free)?
Is the handling of the organism appropriate? (to minimise any stress)
Reliabiliy-test setup
Is the experimental part sufficiently detailed?
Is the study carried out according to a standard protocol/guidance?
If guideline/guidance is available, are validity criteria respected?
Is a negative control present?
If no guideline/guidance is available for the study, is the negative control performing adequately?

Is the application rate/dosing clearly reported?
Is the application type clearly reported?
Is the application timing (date, season, crop growth stage) clearly reported?
Is the treated crop clearly identified?
Is the study location sufficiently identified in geographical terms?
Are the frequency and the duration of sampling/observation clearly identified and appropriate?
Is the sampling/observation technique reported and is it appropriate?
Is the colony (colonies) size appropriate for the study?
Are landscape characteristics sufficiently described and appropriate (e.g. land use, potential dilution)?
Are other crop management procedures (likely to have an impact) sufficiently detailed?
Are weather conditions sufficiently reported?
Is the plot unity of appropriate size?
Is the distance between treatments and control appropriate?
Are conditions in treatments and control comparable at the beginnig of the experiment?
Is the soil typology identified (taxonomy, pH, OC%)?
Is the study methodology consistent with the objective of the study?
Is the presence of other stressors checked and accounted for?
Reliability- Exposure characterisation
Is the test item clearly identified and characterised?
For studies with technical - Is the purity of the a.s. reported?
For studies with formulations - Is the a.s. content specified?
For studies with formulations - Is the effect of coformulant investigated?
Is the exposure duration clearly indicated and appropriate?
Is exposure characterised by analytical measurements (residues or confirmed dose)?
Is "dustiness" of treated seed reported (i.e. Heubach test)?
Reliability- Test design and statistics
Are all measured endpoints/response variables clearly identified in the methodology description?
Is the test sample size/observation unit appropriate?
Is replication appropriate?
Are appropriate statistics available, described and consistent with the initial proposal of the study?
Are appropriate data elaboration presented?
Are raw data available?
Are units correct and consistent within the study?
Is the test design/sampling randomised?
Study conclusions
Is the conclusion of the study supported by the data/results?

e. Experiment type 5 – Residue measurement [soil applications]

Relevance
In case of mixture/formulation: is the test item representative and relevant for the a.s.?
Is the application method relevant for the uses under evaluation?
Is the treated crop relevant for the investigated GAPs?
Is the study location relevant for the investigated GAPs?
Is the application period (season, growth stage) relevant for the investigated GAPs?
Is the matrix analysed relevant for the risk assessment to bees?

Reliability-test organisms
Is/are the tested organism(s) clearly identified (species level or lower)?
Reliabiliy-test setup
Is the experimental part sufficiently detailed?
Is the study carried out according to a standard protocol/guidance?
If guideline/guidance is available, are validity criteria respected?
Is a blank control present?
Is the application **rate** clearly reported?
Is the application **type** clearly reported?
Is the application **timing** (date, season, crop growth stage) clearly reported?
Is the treated crop clearly identified?
Is the study location sufficiently identified in geographical terms?
Are the frequency and the duration of sampling/observation clearly identified and appropriate?
Is the sampling technique reported and is it appropriate?
Are weather conditions sufficiently reported?
Is the sampling time reported (BBCH stage/date)?
Is the soil typology identified (taxonomy, pH, OC%)?
Is the study methodology consistent with the objective of the study?
Reliability- Exposure characterisation
Is the test item clearly identified and characterised?
For studies with technical - Is the purity of the a.s. reported?
For studies with formulations - Is the a.s. content specified?
Was the residue decline sufficiently described?
Is "dustiness" of treated seed reported (i.e. Heubach test)?
Notes (exposure characterisation)
Reliability- Test design and statistics
Is replication appropriate?
Are appropriate statistics available, described and consistent with the initial proposal of the study?
Are appropriate data elaboration presented?
Are raw data available?
Are units correct and consistent within the study?
Is the test design/sampling randomised?
Notes (test design and statistics)
Analytics
Is the analysed matrix clearly identified?
Is the sample storage and stability reported and appropriate?
Is equipment, materials and conditions used clearly described?
Is sample preparation techniques reported?
Is the method for the extraction and purification of the samples clearly described?
Is the analytical procedure clearly reported?
Is the described method appropriate for the intended analysis?

Linearity: is the analytical calibration extended over an appropriate range?
Linearity: is the number of determinations suitable for calibration?
Linearity: are details equation of the calibration line reported?
Linearity: is the overall calibration (linearity) satisfactory?
Accuracy: is mean recovery (+ RSD) reported for each investigated compound?
Accuracy: is mean recovery satisfactory?
Precision: is the RSD reported for each fortification level and for the overall recovery?
Precision: is the RSD < 20% per each fortification level?
LOQ: Is the LOQ clearly reported?
LOQ: are blank values (procedural blanks and untreated samples) < 30% of the LOQ?
Are the representative chromatograms reported?
Were the control samples analysed concurrently to determine any contamination by the analyte of interest or interferences?
Are all performed calculations clear and correct?
Is the analytical sample size appropriate?
Study conclusions
Is the conclusion of the study supported by the data/results?

f. Experiment type 6 – Bee monitoring [seed treatment or granules]

Relevance
Could the endpoint be relevant for the risk assessment according to EFSA (2013)?
In case of mixture/formulation: is the test item representative and relevant for the a.s.?
Is the application method relevant for the uses under evaluation?
Is the monitored area representative for one or more crops relevant for the investigated GAPs?
Is the study location relevant for the investigated GAPs?
Is the application period (season, growth stage) relevant for the investigated GAPs?
Reliability-test organisms
Is/are the monitored organism(s) clearly identified (species level or lower)?
Is the health status of the organisms checked and appropriate (are they disease-free)?
Reliabiliy-test setup
Is the experimental part sufficiently detailed?
Is a reference site present?
Is the application rate/dosing clearly reported?
Is the application type clearly reported?
Is the application timing (date, season, crop growth stage) clearly reported?
Is/are the treated crop(s) clearly identified?
Is the study location sufficiently identified in geographical terms?
Are the frequency and the duration of sampling/observation clearly identified and appropriate?
Is the sampling/observation technique reported and is it appropriate?
Are landscape characteristics sufficiently described (e.g. land use, potential dilution)?
Are other crop management procedures (likely to have an impact) sufficiently detailed?
Are weather conditions sufficiently reported?

Is the distance between treated area and reference area appropriate?
Are conditions in the treated area and in the reference area comparable at the beginning of the monitoring?
Is the soil typology identified (taxonomy, pH, OC%)?
Is the study methodology consistent with the objective of the study?
Is the presence of other stressors checked and accounted for?
Reliability- Exposure characterisation
Is the test item clearly identified and characterised?
For studies with formulations - Is the a.s. content specified?
Is exposure characterised by analytical measurements (residues or confirmed dose)?
Is "dustiness" of treated seed reported (i.e. Heubach test)?
Reliability- Test design and statistics
Are all measured endpoints/response variables clearly identified in the methodology description?
Is the test sample size/observation unit appropriate?
Is replication appropriate?
Are appropriate statistics available, described and consistent with the initial proposal of the study?
Are appropriate data elaboration presented?
Are raw data available?
Are units correct and consistent within the study?
Is the test design/sampling randomised?
Study conclusions
Is the conclusion of the study supported by the data/results?

Appendix B – Title/abstract screening phase and outcome of the full text retrieval process

Appendix B can be found in the online version of this output ('Supporting information' section) http://dx.doi.org/10.2903/j.efsa.2018.1378

Appendix C – Full text screening phase and experiment type classification

Appendix C can be found in the online version of this output ('Supporting information' section) http://dx.doi.org/10.2903/j.efsa.2018.1378

Appendices from D to O

Appendices from D to O are provided as a separate document which can be found in the online version of this output ('Supporting information' section): http://dx.doi.org/10.2903/j.efsa.2018.1378

Appendix D – Study evaluation notes for clothianidin

Appendix E – Study evaluation notes for imidacloprid

Appendix F – Study evaluation notes for thiamethoxam

Appendix G – Study evaluation notes for clothianidin + imidacloprid (no interaction)

Appendix H – Study evaluation notes for clothianidin * imidacloprid (with interaction)

Appendix I – Study evaluation notes for clothianidin * thiamethoxam (with interaction)

Appendix J – Study evaluation notes for clothianidin + thiamethoxam (no interaction)

Appendix K – Study evaluation notes for imidacloprid * thiamethoxam (with interaction)

Appendix L – Study evaluation notes for imidacloprid + thiamethoxam (no interaction)

Appendix M – Study evaluation notes for All+ substances (no interaction)

Appendix N – Study evaluation notes for All* substances (with interaction)

Appendix O – Study evaluation notes for not substance-specific assessments